Journal of Multivariate Analysis **81**, 301–334 (2002)
doi:10.1006/jmva.2001.2009, available online at http://www.idealibrary.com on IDEAL®

Bayesian Object Identification: Variants[1]

Gunter Ritter[2] and María Teresa Gallegos

Universität Passau, D-94 030 Passau, Germany

Received April 19, 1999; published online December 12, 2001

We present a Bayesian theory of object identification. Here, identifying an object means selecting a particular observation from a group of observations (*variants*), this observation (the *regular* variant) being characterized by a distributional model. In this sense, object identification means assigning a given model to one of several observations. Often, it is the statistical model of the regular variant, only, that is known. We study an estimator which relies essentially on this model and not on the characteristics of the “irregular” variants. In particular, we investigate under what [illegible] to the case of irregular variants generated from the regular variant by a Gaussian linear model.

AMS 1991 Subject Classifications: 62C10; 62H30.

Key words and phrases: Bayesian identification; selection of variants; simple selector; exchangeability; detailed balance; reversibility.

1. INTRODUCTION

The subject of this paper is a Bayesian theory of object identification. Suppose that we are given a statistical model and various (in general multivariate) observations; exactly one of the observations belongs to the model. The task is to decide which one it is. The observations may either emanate from different physical objects or they may be different measurements from the same one. In the latter case, the observations foreign to the model may be considered as (deterministic or random) perturbations of the correct observation of the object.

We call the various observations at hand *variants* and the observation belonging to the model the *regular* variant. Two problems arise from the consideration of variants: *selection* of the regular variant from the set of observations and *classification* of an object into one of several classes in the

[1] Research supported by Deutsche Forschungsgemeinschaft Grant Ri477/4.
[2] E-mail: ritter@fmi.uni-passau.de.

0047-259X/01 $35.00

presence of variants of the object. We deal here with the former problem; it is also a building block for the solution of the latter [6, 5] which will be treated in more detail in future communications. An application of the Bayesian paradigm to the present situation immediately shows that it requires the *joint* distribution of all variants, cf. 2.1. This, however, is often unknown. The reason for this lack of information is the fact that irregular variants may be spurious and different trials may even produce different numbers of them. The whole scenario may be too complex and uncertain. We, therefore, focus on a simple estimator that depends mainly on the distribution of the regular variant alone.

Let $x_1, ..., x_b \in E$ be the measurements of b variants $1..b$ in some unknown order, let $f^{\rho}_{Z_1}$ be the density function (with respect to some reference measure ρ on E) of the random features $Z_1 \colon \Omega \to E$ of the regular variant, and let q_h be the prior probability for each position $h \in 1..b$ to be that of the regular variant. We are not given the densities of the other variants, let alone the joint densities of all variants. The task consists in selecting the position of the regular variant. It is tempting to choose an index h for which the product $f^{\rho}_{Z_1}(x_h)\, q_h$ is maximal, i.e., to use the estimator

$$SS^{\rho}(x_1, ..., x_b) = \underset{h}{\operatorname{argmax}}\ f^{\rho}_{Z_1}(x_h)\, q_h.$$

We call such an estimator a *simple (variant) selector*. The simple selector may, however, be grossly misleading and it is the purpose of the present communication to give and study conditions which guarantee its optimality or, in other words, to investigate when the *Bayesian selector* just depends on the reduced set of quantities. We will show that there do exist interesting situations where this is the case, cf. also Discussion 4.7(c). Besides having an algorithmically simple form the simple selector does not require the full statistical model and, thus, admits safer parameter estimation with less noise. We will, however, not deal with the issue of estimating the function $f^{\rho}_{Z_1}$ here.

We see an analogy between simple variant selection and statistical sufficiency. In sufficiency, partial information on the *observations* suffices to optimally identify a model. Simple selectors need partial information on the *model* and we study conditions so that this partial information allows to optimally estimate the regular observation.

Variant selection must not be confused with *classification* (or object *recognition*), *tests of hypotheses*, and *goodness-of-fit tests*. In some sense, variant selection is even converse to classification since, in the former case, several observations compete for one statistical model and in the latter case, several statistical models compete for an observation. A test of hypotheses needs statistical models of both hypotheses and is in this sense similar to classification. Finally, a goodness-of-fit test compares two models with each other, one of them being accessible through realizations, only.

In Section 2 we introduce simple selectors and relate them to the (optimal) Bayesian selector. As to be expected, the latter cannot always be represented as a simple selector with respect to some reference measure, Example 2.4.2. However, in Section 3, we find sufficient conditions for simple selectors to be optimal. In particular, we point out a relationship with exchangeable measures and the equation of detailed balance known from the theory of Markov processes. Finally, in Section 4, we apply the theory of Section 3 to the case of irregular variants arising from the regular one as linear, Gaussian perturbations.

In connection with *segmentation problems* in optical character recognition, the idea of variants has been known for some time as a heuristic method; cf. the survey article [2], pp. 698. In fact, the consideration of variants marks one of the main achievements in this field in the last decades. In the engineering literature, segmentation variants are often called "segmentation hypotheses" but we prefer the term "variant" to "hypothesis" since the latter has a different meaning in statistics. To the best of our knowledge, variant selection in our sense has not yet been analyzed from a mathematical–statistical point of view and, in fact, it seems that the question of its optimality has never been raised. This communication is meant as a first step towards a Bayesian analysis.

Furthermore, we noticed that variants, viewed from their natural level of generalization, have a much broader range of application than just segmentation. The idea may be used, e.g., to handle ambiguities in the process of feature extraction from an object in pattern recognition. A typical example is this: Often, the shape of a geometric object cannot unambiguously be recognized at the low level of image processing. It may then depend on internal features of the object but, usually, these features can only be extracted after the object's shape is recognized. A way out of this vicious circle is the consideration of variants, i.e., the extraction of several feature sets, one for each reasonable shape assignment. By selecting the feature set that fits the model best it is then possible to determine the "true" shape. In this way shape recognition amounts to object identification.

Recently, we successfully applied variants in various contexts to "automatic classification of chromosomes". The correct polarity of a chromosome under a light microscope is not a priori given, a situation giving rise to considering two variants, i.e., one feature set for each polarity. After collecting information at a higher level, the most prospective of them is selected. The resulting "polarity free" classification method reduces the error rate by about 25% [7]. Some methods of automatic chromosome classification require the extraction of longitudinal axes for feature measurement. In the case of a severely bent, badly shaped, or small chromosome the axis (and, hence, the shape) is not easily determined and a way of handling this uncertainty is the simultaneous consideration of various possible axes [8, 9]. Variants thus help to attain the presently worldwide lowest error rate of 0.6% in this field.

General notation. The pair (Ω, P) denotes a probability space and E a Polish state space with its Borel σ-algebra $\mathscr{B}(E)$. Given another measurable space $(F, \mathscr{B}(F))$, a Markov kernel $N: E \times \mathscr{B}(F) \to [0, 1]$, and a measure ρ on E, $\rho \otimes N$ stands for the measure $\rho(dx)\, N(x, dy)$ on $E \times F$. The expression $\nu^{\otimes n}$ denotes the n-fold tensor product on E^n of some distribution ν on E. If μ is any measure on F and φ any measurable mapping from F to another measurable space, μ_φ denotes the induced measure. Given two measures μ, ν on E, the notation $\mu \ll \nu$ means absolute continuity of μ with respect to ν while $\sim$ means their equivalence, i.e., mutual absolute continuity. The symbol $N(m, V)$ denotes the (uni- or multi-variate) normal distribution with expectation m and variance V; $n(m, V)$ stands for its density function with respect to Lebesgue measure. For details on measure spaces and kernels the interested reader is referred to Bauer [1].

Let $b \in \mathbf{N}$. The symbol $\mathscr{S}_b$ denotes the symmetric group of b elements and $\mathscr{S}_{b,h}$ the set of all permutations $\pi \in \mathscr{S}_b$ such that $\pi(h) = 1$. The notation $(j, k) \in \mathscr{S}_b$ stands for the transposition of $j, k \in 1..b$. Given a random variable $Z: \Omega \to E^b$ and $\pi \in \mathscr{S}_b$, Z_π stands for the random variable $(Z_{\pi(1)}, ..., Z_{\pi(b)})$ and $P_{Z_\pi} = P_{(Z_\pi)}$. A hat on top of an index set $I \subseteq 1..b$ indicates missing entries in an array: Given $\mathbf{x} = (x_1, ..., x_b) \in E^b$ we write $\mathbf{x}_{\hat{I}} = (x_i)_{i \notin I} \in E^{\complement I}$; in particular,

$$\mathbf{x}_{\hat{h}} = (x_1, ..., x_{h-1}, x_{h+1}, ..., x_b) \in E^{b-1}.$$

Given a real-valued function $g: \mathscr{F} \to \mathbf{R}$ on a finite set $\mathscr{F}$, $\operatorname{argmax}_{h \in \mathscr{F}} g(h)$ ($\operatorname{argmin}_{h \in \mathscr{F}} g(h)$) denotes the set of its maximal (minimal) arguments.

2. BAYESIAN AND SIMPLE SELECTION OF THE REGULAR VARIANT

2.1. *The Statistical Model*

We first cast the selection problem in a Bayesian framework. Let $Z_i: (\Omega, P) \to E$, $i \in 1..b$, be $b \geqslant 1$ random variants, Z_1 being the regular one. We observe a realization $\mathbf{x} = (x_1, ..., x_b) \in E^b$ of $X = (X_1, ..., X_b) = (Z_{T(1)}, ..., Z_{T(b)}) = Z_T$, a random permutation $T: \Omega \to \mathscr{S}_b$ of the b-tuple

$Z=(Z_1, \ldots, Z_b)$. Our task is to estimate the unknown random position $H\colon \Omega \to 1..b$ of the regular variant 1, i.e., the position H so that x_H emanates from Z_1. Clearly, we have $T(H)=1$, $H=T^{-1}(1)$, and the assertions $T(h)=1$ and $H=h$ are synonymous. We assume that the random permutation T is independent of Z.

The related statistical model is the quadruple

$$(X, (P_{Z_\pi})_{\pi \in \mathscr{S}_b}, \mathscr{D}, G).$$

Here, the symmetric group $\mathscr{S}_b$ is the parameter set, the interval $\mathscr{D}=1..b$ is the decision set, and we define our gain function $G\colon \mathscr{S}_b \times (1..b) \to \mathbf{R}$ as

$$G(\pi, h) = \begin{cases} 1, & \text{if} \quad \pi(h)=1, \\ 0, & \text{otherwise.} \end{cases}$$

For all $h \in 1..b$, let $q_h := P[T(h)=1]$ denote the prior probability of the regular variant to occupy site h. Without loss of generality we assume $q_h > 0$ for all $h \in 1..b$. We will call any estimator $S\colon E^b \to 1..b$ of the regular variant a *selector*.

The *Bayesian selector BS* for the statistical model at hand is the subset of $1..b$ defined by

$$\text{(1)} \qquad BS(\mathbf{x}) = \underset{h \in 1..b}{\operatorname{argmax}}\ \mathrm{E}[\mathrm{G(T, h)} \mid \mathrm{X}=\mathbf{x}] = \underset{h \in 1..b}{\operatorname{argmax}}\ \mathrm{P}[\mathrm{T(h)}=1 \mid \mathrm{Z_T}=\mathbf{x}].$$

This set is nonempty and uniquely defined for P_{Z_T}-a.a. $\mathbf{x} \in E^b$. Let μ be some σ-finite measure on E^b such that $P_{Z_T} \ll \mu$. By Bayes' formula, $P[T(h)=1 \mid Z_T=\mathbf{x}]$ equals the density

$$\text{(2)} \qquad \frac{P[Z_T \in d\mathbf{x}, T(h)=1]}{P[Z_T \in d\mathbf{x}]} = \frac{P[Z_T \in d\mathbf{x}, T(h)=1]}{\mu(d\mathbf{x})} \Bigg/ \frac{P[Z_T \in d\mathbf{x}]}{\mu(d\mathbf{x})}.$$

This implies for P_{Z_T}-a.a. $\mathbf{x}$

$$\text{(3)} \qquad BS(\mathbf{x}) = \underset{h \in 1..b}{\operatorname{argmax}}\ \frac{\mathrm{P}[\mathrm{Z_T} \in \mathrm{d}\mathbf{x}, \mathrm{T(h)}=1]}{\mu(\mathrm{d}\mathbf{x})}.$$

This selector needs information on the *joint* distribution of *all* variants. However, complete knowledge of the characteristics of the irregular variants, let alone the joint distribution of all variants, will often not be available. This is one reason why we introduce another selector, the *simple selector*, which we next define in a more formal way.

2.2. *Notation*

Let ρ be any σ–finite *reference measure* on E such that $P_{Z_1} \ll \rho$ and let $f^{\rho}_{Z_1}$ denote some version of the density $dP_{Z_1}/d\rho$. For all $\mathbf{x} \in E^b$, let

$$\mathscr{P}(\mathbf{x}) = \{h \in \mathscr{D} / q_h f^{\rho}_{Z_1}(x_h) > 0\}.$$

Let us note that $\mathscr{P}(\mathbf{x})$ is nonempty for P_{Z_T}-a.a. $\mathbf{x} \in E^b$. Indeed, let $N^{(l)} = \{\mathbf{x} \in E^b / q_l f^{\rho}_{Z_1}(x_l) = 0\}$. This set is of the form $N^{(l)} = E^{l-1} \times N_l \times E^{b-l}$ where $N_l = \{x \in E / q_l f^{\rho}_{Z_1}(x) = 0\}$ and we have $\{\mathbf{x} \in E^b / \mathscr{P}(\mathbf{x}) = \varnothing\} = \bigcap_l N^{(l)}$. Now,

$$P[Z_T \in \bigcap_l N^{(l)}] = \sum_h P[Z_T \in \bigcap_l N^{(l)}, T(h) = 1] \leqslant \sum_h P[Z_T \in N^{(h)}, T(h) = 1]$$

$$= \sum_h P[Z_1 \in N_h, T(h) = 1] = \sum_h q_h P[Z_1 \in N_h] = 0.$$

If Z_1 is discrete then $f^{\rho}_{Z_1}(x) = P[Z_1 = x]/\rho(x)$ and $\mathscr{P}(\mathbf{x}) = \{h \in \mathscr{D} / q_h P[Z_1 = x_h] > 0\}$, the set of all h such that x_h belongs to the support of P_{Z_1} (recall $q_h > 0$).

Contrary to the Bayesian selector, which does not depend on the measure μ on E^b, the following simple selector depends on the reference measure ρ.

2.3. *Definition: Simple Selector*

If $\mathscr{P}(\mathbf{x}) \neq \varnothing$, we define the *simple selector* associated with the reference measure ρ as

$$SS^{\rho}(\mathbf{x}) = \underset{\mathrm{h} \in \mathscr{P}(\mathbf{x})}{\operatorname{argmax}} \; \mathrm{f}^{\rho}_{Z_1}(\mathrm{x_h})\, \mathrm{q_h}.$$

The simple selector SS^{ρ} is defined P_{Z_T}-a.s. on E^b since $\mathscr{P}(\mathbf{x})$ is nonempty for P_{Z_T}-a.a. $\mathbf{x}$. The reference measure ρ may be replaced with its absolutely continuous part with respect to P_{Z_1}, $P_{Z_1}/f^{\rho}_{Z_1}$, without changing the simple selector. It may, thus, be assumed to be equivalent to P_{Z_1}. The main objective of this paper is to reveal reference measures ρ such that SS^{ρ} is optimal, i.e., equals the Bayesian selector, $SS^{\rho}(\mathbf{x}) = BS(\mathbf{x})$ for P_{Z_T}-a.a. $\mathbf{x}$. We first take a look at some instructive examples.

2.4. *Examples*

2.4.1. *Two normally-distributed variants.* By way of illustration let us consider the case of $b = 2$ univariate variants Z_1 and Z_2. We assume that the joint vector (Z_1, Z_2): $(\Omega, P) \to \mathbf{R}^2$ is normally distributed $N((0, m), V)$ with $m > 0$ and $V := \left(\begin{smallmatrix} 1 & k \\ k & 1 \end{smallmatrix}\right)$, $k \in]-1, 1[$. Thus, both variants have variance 1, the regular variant being standard normal; the parameter m is the expectation of the second variant and $k \in]-1, 1[$ is their coefficient of correlation. Moreover, we assume $q_1 = q_2 = 1/2$.

We study the Bayesian selector and the two simple selectors associated with Lebesgue measure λ and the Gaussian measure $\nu = N(a, 1)$, $a > 0$, respectively, as reference measures. We prove the following assertions, valid for $\mathbf{x} = (x_1, x_2) \in \mathbf{R}^2$:

(a) $BS(\mathbf{x}) = \{h \in 1..2 / x_h \leqslant x_{3-h}\}$;

(b) (i) $SS^\lambda(\mathbf{x}) = \{h \in 1..2 / x_h^2 \leqslant x_{3-h}^2\}$,

(ii) $SS^\lambda(\mathbf{x}) \subseteq BS(\mathbf{x})$ if and only if $x_1 + x_2 > 0$ or $x_1 = x_2$ and, in this case, we have $SS^\lambda(\mathbf{x}) = BS(\mathbf{x})$;

(c) (i) $SS^\nu(\mathbf{x}) = \{h \in 1..2 / x_h \leqslant x_{3-h}\}$,

(ii) $SS^\nu = BS$.

Proof. (a) The Bayesian selector for $b = 2$ variants is a standard discriminant rule between two classes, the identity permutation and the transposition. Both classes are normally distributed: the identity according to $(Z_1, Z_2) \sim N((0, m), V)$ and the transposition according to $(Z_2, Z_1) \sim N((m, 0), V)$. Hence, we have $h \in BS(\mathbf{x})$ if and only if

$$(x_h, x_{3-h} - m)\, V^{-1} (x_h, x_{3-h} - m)^T \leqslant (x_h - m, x_{3-h})\, V^{-1} (x_h - m, x_{3-h})^T.$$

Now, $V^{-1} = 1/(1-k^2) \left(\begin{smallmatrix} 1 & -k \\ -k & 1 \end{smallmatrix}\right)$ and a simple computation shows Claim (a).

(b) According to 2.3, we have $h \in SS^\lambda(\mathbf{x})$ if and only if $f^\lambda_{Z_1}(x_h) \geqslant f^\lambda_{Z_1}(x_{3-h})$, i.e.,

$$n(0, 1)(x_h) \geqslant n(0, 1)(x_{3-h}).$$

This is plainly equivalent to having $x_h^2 \leqslant x_{3-h}^2$.

By (a) and (i) we have $SS^\lambda(\mathbf{x}) \subseteq BS(\mathbf{x})$ if and only if, for all h, $(x_h - x_{3-h})(x_1 + x_2) \leqslant 0$ implies $x_h - x_{3-h} \leqslant 0$. The first half of Claim (ii) now follows from considering the three cases $x_1 + x_2\ >,\ <$, and $= 0$ and the second half is immediate.

(c) In this case, $f^\nu_{Z_1}(x) = \frac{n(0, 1)(x)}{n(a, 1)(x)} = c \cdot e^{-ax}$ with a constant $c > 0$. Recalling $a > 0$, we have $h \in SS^\nu(\mathbf{x})$ if and only if $x_h \leqslant x_{3-h}$ which proves Claims (i) and (ii). ∎

Part (b)(ii) shows that, no matter how large m is, i.e., how well the distributions are separated, the simple selector may not be optimal. On the other hand, Part (c) shows that, in the present normal case, there always exist reference measures ρ such that the simple selector is optimal for all observations $\mathbf{x}$! We will see in Section 4 that this is no coincidence. The following example shows that the situation may be worse.

2.4.2. *Functional dependence.* Let us show that there need not exist ρ such that the simple selector is Bayesian given full information on Z_1 and Z_2 even if Z_2 is a function of the regular variant Z_1.

Let $E = \mathbb{Z}/3\mathbb{Z}$, let $b = 2$, let Z_1 be uniformly distributed on E, and let $Z_2 = Z_1 + 1 \pmod 3$. The prior probabilities are $q_1 = q_2 = 1/2$. The only possible observations are (0, 1), (1, 0), (1, 2), (2, 1), (2, 0), and (0, 2). The Bayesian selector is always correct. Now, for any reference measure ρ on E, $\rho > 0$ on E, we have

$$SS^\rho(\mathbf{x}) = \operatorname*{argmin}_{\mathrm{h} \in 1..2} \rho(\mathrm{x_h});$$

this establishes a preference ordering on E and, hence, the simple selector SS^ρ cannot always be contained in the Bayesian selector.

2.5. *Error Probability*

It is easy to compare the error probabilities of the Bayesian selector and the simple selector in the case of two variants and a continuous distribution P_Z. In this case the position of the regular variant uniquely defines the permutation. Let tr be the transposition $\in \mathcal{S}_2$. The Bayesian selector chooses Position 1 if $f_Z(X) > f_{Z_{tr}}(X)$ and commits an error when the true position is 2, i.e., $T = tr$; it chooses Position 2 if $f_Z(X) < f_{Z_{tr}}(X)$ and commits an error when $T = id$. Thus, its error probability is

$$\begin{aligned} err_{BS} &= P[T = tr,\, f_Z(X) > f_{Z_{tr}}(X)] + P[T = id,\, f_Z(X) < f_{Z_{tr}}(X)] \\ &= P[T = tr,\, f_Z(Z_{tr}) > f_Z(Z)] + P[T = id,\, f_Z(Z) < f_{Z_{tr}}(Z)]. \end{aligned}$$

Now, $f_Z(z_{tr}) = f_{Z_{tr}}(z)$ and, hence, err_{BS} equals

$$\begin{aligned} &P[T = tr,\, f_{Z_{tr}}(Z) > f_Z(Z)] + P[T = id,\, f_Z(Z) < f_{Z_{tr}}(Z)] \\ &\quad = P[f_Z(Z) < f_{Z_{tr}}(Z)]. \end{aligned}$$

Denoting $A_{BS} = \{z \in E^2 / f_Z(z) < f_{Z_{tr}}(z)\}$, we have $err_{BS} = P[Z \in A_{BS}]$.

The simple selector chooses Position 1 if $f^\rho_{Z_1}(X_1) > f^\rho_{Z_1}(X_2)$ which is an error when $T = tr$; it chooses Position 2 if $f^\rho_{Z_1}(X_2) > f^\rho_{Z_1}(X_1)$ committing an error when $T = id$. Thus, its error probability is

$$\begin{aligned} err_{SS^\rho} &= P[T = tr,\, f^\rho_{Z_1}(X_1) > f^\rho_{Z_1}(X_2)] + P[T = id,\, f^\rho_{Z_1}(X_2) > f^\rho_{Z_1}(X_1)] \\ &= P[T = tr,\, f^\rho_{Z_1}(Z_2) > f^\rho_{Z_1}(Z_1)] + P[T = id,\, f^\rho_{Z_1}(Z_2) > f^\rho_{Z_1}(Z_1)] \\ &= P[f^\rho_{Z_1}(Z_1) < f^\rho_{Z_1}(Z_2)]. \end{aligned}$$

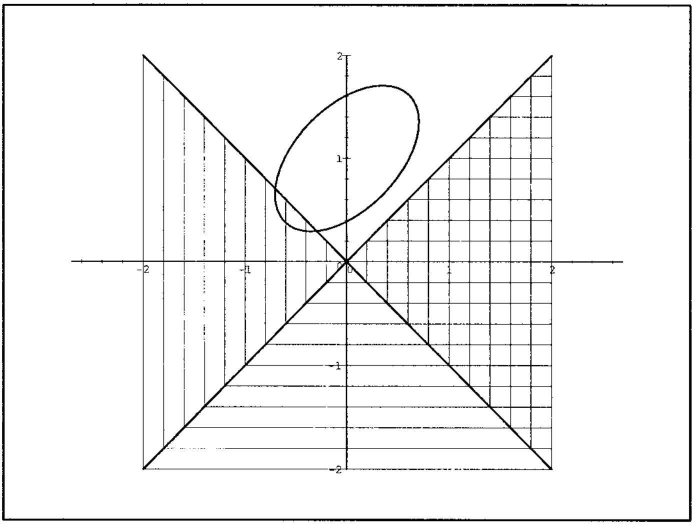

FIG. 1. The error sets A_{BS} and A_{SS^λ} for the selectors BS and SS^λ of Example 2.4.1. Horizontal hatching, A_{BS}; vertical, A_{SS^λ}.

Letting $A_{SS^\rho} = \{(z_1, z_2) \in E^2 / f^\rho_{Z_1}(z_1) \leqslant f^\rho_{Z_1}(z_2)\}$ we have $err_{SS^\rho} = P[Z \in A_{SS^\rho}]$. A graphical illustration is presented in Fig. 1.

3. SUFFICIENT CONDITIONS FOR OPTIMALITY

In this section, we derive general sufficient conditions for the Simple Selector to be Bayesian. We need some preliminaries. Conditioning on the event $Z_1 = x$ is defined for P_{Z_1}-a.a. $x \in E$, only. Therefore, the following definitions need some care if ρ is not equivalent to P_{Z_1}.

3.1. *Notation and Remarks*

(a) The space E being Polish there is an extension of the conditional distribution $P[Z_{\hat{1}} \in d\mathbf{y} \mid Z_1 = x]$ to a Markovian kernel K: $E \times \mathscr{B}(E^{b-1}) \to [0, 1]$, i.e.,

$$K(x, d\mathbf{y}) = P[Z_{\hat{1}} \in d\mathbf{y} \mid Z_1 = x] \tag{4}$$

for P_{Z_1}-a.a. $x \in E$. Let μ be a σ-finite measure on E^b such that

$$\rho(dx_h)\, K(x_h, d(\mathbf{x}_{\pi^{-1}})_{\hat{1}}) \ll \mu(d\mathbf{x}) \tag{5}$$

for all $\pi \in \mathscr{S}_b$ and all $h \in 1..b$ such that $\pi(h) = 1$. Note that any such measure satisfies $P_{Z_\pi} \ll \mu$ for all $\pi \in \mathscr{S}_b$ and hence $P_{Z_T} \ll \mu$. The measure

$$\mu(d\mathbf{x}) = \sum_{h=1}^{b} \sum_{\pi:\, \pi(h)=1} \rho(dx_h)\, K(x_h, d(\mathbf{x}_{\pi^{-1}})_{\hat{1}}) \tag{6}$$

$$= \sum_{\pi \in \mathscr{S}_b} \rho(dx_{\pi^{-1}(1)})\, K(x_{\pi^{-1}(1)}, d(\mathbf{x}_{\pi^{-1}})_{\hat{1}}),$$

e.g., has the property (5).

(b) Central tools for our analysis are the Markov kernels K_h: $E \times \mathscr{B}(E^{b-1}) \to [0, 1]$ defined by

$$K_h(x_h, d\mathbf{x}_{\hat{h}}) := \frac{1}{q_h} \sum_{\pi:\, \pi(h)=1} K(x_h, d(\mathbf{x}_{\pi^{-1}})_{\hat{1}})\, P[T = \pi], \qquad h \in 1..b. \tag{7}$$

It follows from (5) that $\rho \otimes K_h \ll \mu$. For each $h \in 1..b$, let us choose a version $D^\rho(\cdot, h)$ of the μ-density $d(\rho \otimes K_h)/d\mu$,

$$D^\rho(\mathbf{x}, h) = \frac{\rho(dx_h)\, K_h(x_h, d\mathbf{x}_{\hat{h}})}{\mu(d\mathbf{x})}.$$

This density function depends on K, ρ, μ, and h; it is unique up to μ-equivalence.

We will need the following measure-theoretic lemma.

3.2. *Lemma*

Let L: $E \times \mathscr{B}(F) \to [0, 1]$ be a Markov kernel, let β and μ be σ-finite measures on E and $E \times F$, respectively, and let α be a probability measure on E. If $\alpha \ll \beta$ and $\beta \otimes L \ll \mu$ then $\alpha \otimes L \ll \mu$ and, μ-a.s.,

$$\frac{d\alpha \otimes L}{d\mu} = \frac{d\alpha}{d\beta} \cdot \frac{d\beta \otimes L}{d\mu}. \tag{8}$$

Proof. Integrating a measurable function $f: E \times F \to \mathbf{R}_+$, using the identity

$$\int g \frac{d\kappa}{d\nu}\, d\nu = \int g\, d\kappa, \tag{9}$$

and using Fubini's theorem twice we may compute

$$\int_{E\times F} \mu(dx, dy) \frac{\alpha(dx)}{\beta(dx)} \cdot \frac{\beta \otimes L(dx, dy)}{\mu(dx, dy)} f(x, y)$$

$$= \int_{E\times F} \beta \otimes L(dx, dy) \frac{\alpha(dx)}{\beta(dx)} f(x, y)$$

$$= \int_E \beta(dx) \frac{\alpha(dx)}{\beta(dx)} \int_F L(x, dy)\, f(x, y) = \int_E \alpha(dx) \int_F L(x, dy)\, f(x, y)$$

$$= \int_{E\times F} \alpha \otimes L(dx, dy)\, f(x, y).$$

This being true for all such functions f, the lemma follows. ∎

3.3. *Lemma*

(a) For all $h \in 1..b$ and P_{Z_1}-a.a. $x \in E$, we have

$$K_h(x, d\mathbf{y}) = P[(Z_T)_{\hat{h}} \in d\mathbf{y} \mid Z_1 = x, T(h) = 1].$$

(b) For all $h \in 1..b$, we have

$$P[Z_T \in d\mathbf{x} \mid T(h) = 1] = P[Z_1 \in dx_h]\, K_h(x_h, d\mathbf{x}_{\hat{h}}).$$

Proof. (a) By independence of Z and T, we obtain from (7) and (4)

$$K_h(x_h, d\mathbf{x}_{\hat{h}}) = \frac{1}{q_h} \sum_{\pi:\, \pi(h)=1} P[Z_{\hat{1}} \in d(\mathbf{x}_{\pi^{-1}})_{\hat{1}} \mid Z_1 = x_h]\, P[T = \pi]$$

$$= \frac{1}{q_h} \sum_{\pi:\, \pi(h)=1} P[(Z_\pi)_{\hat{h}} \in d\mathbf{x}_{\hat{h}} \mid Z_1 = x_h]\, P[T = \pi]$$

$$= \frac{1}{q_h} \sum_{\pi:\, \pi(h)=1} P[(Z_\pi)_{\hat{h}} \in d\mathbf{x}_{\hat{h}} \mid Z_1 = x_h]\, P[T = \pi \mid Z_1 = x_h]$$

$$= \frac{1}{q_h} P[(Z_T)_{\hat{h}} \in d\mathbf{x}_{\hat{h}}, T(h) = 1 \mid Z_1 = x_h]$$

$$= P[(Z_T)_{\hat{h}} \in d\mathbf{x}_{\hat{h}} \mid Z_1 = x_h, T(h) = 1].$$

(b) Use independence of Z and T and (a) in order to compute for $h \in 1..b$

(10) $$P[Z_T \in d\mathbf{x}, T(h) = 1]$$
$$= P[(Z_T)_{\hat{h}} \in d\mathbf{x}_{\hat{h}}, Z_1 \in dx_h, T(h) = 1]$$
$$= P[(Z_T)_{\hat{h}} \in d\mathbf{x}_{\hat{h}} \mid Z_1 = x_h, T(h) = 1]\; P[Z_1 \in dx_h]\; P[T(h) = 1]$$
$$= P[Z_1 \in dx_h]\; K_h(x_h, d\mathbf{x}_{\hat{h}})\; P[T(h) = 1].$$

Hence, Part (b) follows. ▌

We next formulate two conditions which turn out to be sufficient for optimality of the simple selector.

Conditions $(\mathscr{H}_{\mathbf{x}}^{\rho})$ *and* $(\mathscr{C}_{\mathbf{x}}^{\rho})$.

$(\mathscr{H}_{\mathbf{x}}^{\rho})$ $SS^{\rho}(\mathbf{x}) \subseteq \operatorname{argmax}_{h \in \mathscr{P}(\mathbf{x})} D^{\rho}(\mathbf{x}, h)$;

$(\mathscr{C}_{\mathbf{x}}^{\rho})$ the function $D^{\rho}(\mathbf{x}, \cdot)$ is constant on the set $\mathscr{P}(\mathbf{x})$.

3.4. *Theorem*

(a) Condition $(\mathscr{C}_{\mathbf{x}}^{\rho})$ implies $(\mathscr{H}_{\mathbf{x}}^{\rho})$.

(b) For P_{Z_T}-a.a. $\mathbf{x} \in E^b$, Condition $(\mathscr{H}_{\mathbf{x}}^{\rho})$ (or $(\mathscr{C}_{\mathbf{x}}^{\rho})$) implies $SS^{\rho}(\mathbf{x}) = BS(\mathbf{x})$.

Proof. (a) If $(\mathscr{C}_{\mathbf{x}}^{\rho})$ is satisfied then

$$SS^{\rho}(\mathbf{x}) \subseteq \mathscr{P}(\mathbf{x}) = \operatorname*{argmax}_{\mathrm{h} \in \mathscr{P}(\mathbf{x})} \mathrm{D}^{\rho}(\mathbf{x}, \mathrm{h}).$$

(b) By Lemma 3.3(b), we have

$$P[Z_T \in d\mathbf{x}, T(h) = 1] = P[Z_1 \in dx_h]\; K_h(x_h, d\mathbf{x}_{\hat{h}})\; q_h.$$

By (5), Lemma 3.2 is applicable to $F = E^{b-1}$, $L = K_h$, $\alpha = P_{Z_1}$, and $\beta = \rho$. Hence, for μ-a.a. $\mathbf{x} \in E^b$ and for all $h \in \mathscr{P}(\mathbf{x})$, we have

(11) $$\frac{P[Z_T \in d\mathbf{x}, T(h) = 1]}{\mu(d\mathbf{x})} = \frac{P[Z_1 \in dx_h]\; K_h(x_h, d\mathbf{x}_{\hat{h}})}{\mu(d\mathbf{x})}\, q_h$$
$$= \frac{\rho(dx_h)\; K_h(x_h, d\mathbf{x}_{\hat{h}})}{\mu(d\mathbf{x})}\, f_{Z_1}^{\rho}(x_h)\, q_h$$
$$= D^{\rho}(\mathbf{x}, h)\, f_{Z_1}^{\rho}(x_h)\, q_h.$$

This representation, together with (3) and Condition ($\mathscr{H}_{\mathbf{x}}^{\rho}$), shows the inclusion $SS^{\rho}(\mathbf{x}) \subseteq BS(\mathbf{x})$ for P_{Z_T}-a.a. $\mathbf{x} \in E^b$. (Note that both BS and SS^{ρ} are defined P_{Z_T}-a.s..)

For the converse, first note that for P_{Z_T}-a.a. $\mathbf{x} \in E^b$ there exists $h \in \mathscr{P}(\mathbf{x})$ such that $D^{\rho}(\mathbf{x}, h) > 0$. Indeed, by (11), we have

$$\frac{P_{Z_T}(d\mathbf{x})}{\mu(d\mathbf{x})} = \sum_h q_h f_{Z_1}^{\rho}(x_h)\, D^{\rho}(\mathbf{x}, h)$$

and, hence,

$$\begin{aligned} &P_{Z_T}\{\mathbf{x} \in E^b / D^{\rho}(\mathbf{x}, h) = 0 \text{ for all } h \in \mathscr{P}(\mathbf{x})\} \\ &\quad = P_{Z_T}\{\mathbf{x} \in E^b / q_h f_{Z_1}^{\rho}(x_h)\, D^{\rho}(\mathbf{x}, h) = 0 \text{ for all } h \in 1..b\} \\ &\quad = P_{Z_T}\left\{\mathbf{x} \in E^b \middle/ \sum_{h=1}^{b} q_h f_{Z_1}^{\rho}(x_h)\, D^{\rho}(\mathbf{x}, h) = 0\right\} \\ &\quad = 0. \end{aligned}$$

Now, the function $h \mapsto D^{\rho}(\mathbf{x}, h)\, f_{Z_1}^{\rho}(x_h)\, q_h$ has all its maxima in the (nonempty) set $\mathscr{P}(\mathbf{x})$ since they are strictly positive. By ($\mathscr{H}_{\mathbf{x}}^{\rho}$), the set of these maxima coincides with $SS^{\rho}(\mathbf{x})$. Thus, the inclusion $BS(\mathbf{x}) \subseteq SS^{\rho}(\mathbf{x})$ follows from (3) and (11). ∎

3.5. *Examples*

For the sake of illustration we take a look at two examples. The first one is elementary.

3.5.1. *Disjoint variants.* If the random variants Z_i, $i \in \mathscr{P}(\mathbf{x})$, are pairwise disjoint, i.e., if their distributions are mutually singular then we have ($\mathscr{C}_{\mathbf{x}}^{\rho}$) with respect to any reference measure ρ such that $P_{Z_1} \ll \rho$. Indeed, for any observation $\mathbf{x} \in E^b$ the set $\mathscr{P}(\mathbf{x}) = \{h \in \mathscr{D} / q_h f_{Z_1}^{\rho}(x_h) > 0\}$ contains at most one element, hence the claim.

Let us next resume Example 2.4.1 with its notation.

3.5.2. *Two normally distributed variants.*

(a) Condition ($\mathscr{H}_{\mathbf{x}}^{\lambda}$) is satisfied if and only if $0 < x_1 + x_2 \leqslant \frac{2m}{1-k}$ or $x_1 = x_2$.

(b) Condition ($\mathscr{H}_{\mathbf{x}}^{\nu}$) is satisfied if and only if $a \leqslant \frac{m}{1-k}$ or $x_1 = x_2$.

Proof. In order to prove Claim (a) we put $\mu = \lambda^2 =$ the two-dimensional Lebesgue measure and note that $\mathscr{P}(\mathbf{x}) = \mathscr{D}$ for all $\mathbf{x} \in \mathbb{R}^2$. Using [4], Theorem 3.2.4, we first determine the right side of Condition ($\mathscr{H}_{\mathbf{x}}^{\lambda}$).

$$\underset{h}{\operatorname{argmax}} \frac{\mathrm{P}[\mathrm{Z}_2 \in \mathrm{dx}_{3-\mathrm{h}} \mid \mathrm{Z}_1 = \mathrm{x}_\mathrm{h}]\, \lambda(\mathrm{dx}_\mathrm{h})}{\mu(\mathrm{d}\mathbf{x})}$$
$$= \underset{h}{\operatorname{argmax}}\ \mathrm{n}(\mathrm{m}+\mathrm{kx}_\mathrm{h}, 1-\mathrm{k}^2)(\mathrm{x}_{3-\mathrm{h}})$$
$$= \underset{h}{\operatorname{argmin}}\ (\mathrm{x}_{3-\mathrm{h}} - \mathrm{m} - \mathrm{kx}_\mathrm{h})^2.$$

This is the set of all sites h such that

$$(x_{3-h} - m - kx_h)^2 - (x_h - m - kx_{3-h})^2 \leqslant 0$$

or, equivalently,

$$\left((x_1 + x_2) - \frac{2m}{1-k}\right)(x_h - x_{3-h}) \geqslant 0.$$

A comparison with 2.4.1(b)(i) shows Claim (a).

In order to prove (b) we use again the formula above and deal first with the right side of Condition $(\mathscr{H}_\mathbf{x}^\nu)$.

$$\underset{h}{\operatorname{argmax}} \frac{\mathrm{P}[\mathrm{Z}_2 \in \mathrm{dx}_{3-\mathrm{h}} \mid \mathrm{Z}_1 = \mathrm{x}_\mathrm{h}]\, \nu(\mathrm{dx}_\mathrm{h})}{\mu(\mathrm{d}\mathbf{x})}$$
$$= \underset{h}{\operatorname{argmax}}\ \mathrm{n}(\mathrm{m}+\mathrm{kx}_\mathrm{h}, 1-\mathrm{k}^2)(\mathrm{x}_{3-\mathrm{h}})\ \mathrm{n}(\mathrm{a}, 1)(\mathrm{x}_\mathrm{h})$$
$$= \underset{h}{\operatorname{argmin}}\ (\mathrm{x}_{3-\mathrm{h}} - \mathrm{m} - \mathrm{kx}_\mathrm{h})^2 + (1-\mathrm{k}^2)(\mathrm{x}_\mathrm{h} - \mathrm{a})^2.$$

This is the set of all sites h such that

$$(x_{3-h} - m - kx_h)^2 + (1-k^2)(x_h - a)^2 \leqslant (x_h - m - kx_{3-h})^2 + (1-k^2)(x_{3-h} - a)^2$$

or, equivalently,

$$\left(\frac{m}{1-k} - a\right)(x_h - x_{3-h}) \leqslant 0.$$

A comparison with 2.4.1(c)(i) shows Claim (b). ▮

This example shows, among other things, that Condition $(\mathscr{H}_\mathbf{x}^\rho)$ is not necessary for optimality of a simple selector SS^ρ. In Case (a) the set of $\mathbf{x}$'s such that $(\mathscr{H}_\mathbf{x}^\lambda)$ is satisfied is a strict subset of the set where SS^ρ and BS are equal, cf. 2.4.1(b)(ii); in Case (b) it coincides with this set (i.e., the whole plane) if a is small enough.

In the following, our aim is to study the Condition $(\mathscr{C}_{\mathbf{x}}^{\rho})$; we will formulate a number of criteria equivalent with and conditions sufficient for its validity. By Theorem 3.4(b), each of them entails optimality of the simple selector.

3.6. *Theorem*

(a) The following are equivalent.

(α) There exists $\rho \gg P_{Z_1}$ such that Condition $(\mathscr{C}_{\mathbf{x}}^{\rho})$ is satisfied for μ-a.a. $\mathbf{x}$;

(β) there exists a P_{Z_1}-a.e. strictly positive, measurable function $f\colon E \to \mathbb{R}$ such that

(i) $f(x_h)\, P[Z_T \in d\mathbf{x} \mid T(k) = 1] = f(x_k)\, P[Z_T \in d\mathbf{x} \mid T(h) = 1]$ for all $h, k \in 1..b$;

(γ) there exists a P_{Z_1}-a.e. strictly positive, measurable function $f\colon E \to \mathbb{R}$ such that

(ii) $f(x_h)\, P[Z_T \in d\mathbf{x} \mid T(1) = 1] = f(x_1)\, P[Z_T \in d\mathbf{x} \mid T(h) = 1]$ for all $h \in 1..b$;

(δ) there exists a σ-finite measure ν on E^b such that, for all $h \in 1..b$, $P[Z_T \in \cdot \mid T(h) = 1] \ll \nu$ and the density function

$$P[Z_T \in d\mathbf{x} \mid T(h) = 1]/\nu(d\mathbf{x}) = g(x_h)$$

depends on x_h alone.

(b) In this case, we have $f = g = f_{Z_1}^{\rho}$.

Proof. Let $E_h = \{\mathbf{x} \in E^b / f(x_h) > 0\}$, $h \in 1..b$, where f is given by the context.

Assume first (α) and put $f = f_{Z_1}^{\rho}$. For any measurable subset $B \subseteq E^b$ and any $h, k \in 1..b$, we deduce from $(\mathscr{C}_{\mathbf{x}}^{\rho})$

$$\begin{aligned}\int_B f(x_h)\, P[Z_1 \in dx_k]\, K_k(x_k, d\mathbf{x}_{\hat{k}}) &= \int_{B \cap E_h \cap E_k} f(x_h)\, f(x_k)\, \rho(dx_k)\, K_k(x_k, d\mathbf{x}_{\hat{k}}) \\ &= \int_{B \cap E_h \cap E_k} f(x_k)\, f(x_h)\, \rho(dx_h)\, K_h(x_h, d\mathbf{x}_{\hat{h}}) \\ &= \int_B f(x_k)\, P[Z_1 \in dx_h]\, K_h(x_h, d\mathbf{x}_{\hat{h}}).\end{aligned}$$

Statement (i) now follows from 3.3(b). This proves (β) with $f = f_{Z_1}^{\rho}$.

Assume next (β). By assumption and by independence of Z and T,

$$P[Z_T \in E_h \mid T(h) = 1] = P_{Z_1}[f > 0] = 1. \tag{12}$$

Thus, the expression

$$\nu_h(d\mathbf{x}) := \frac{1}{f(x_h)} P[Z_T \in d\mathbf{x} \mid T(h) = 1]$$

defines a σ-finite measure ν_h on E_h and it follows from (i) that $\nu_h = \nu_k$ on $E_h \cap E_k$. Therefore, there is a joint extension ν of all measures ν_h to E^b. Using again (12), we see that the ν-density of $P[Z_T \in d\mathbf{x} \mid Th = 1]$ equals f both on E_h and off E_h, where both sides vanish. This is (δ) with $g = f$.

We next show that (δ) implies (γ). Integrating the equality

$$P[Z_T \in d\mathbf{x} \mid T(h) = 1] = g(x_h)\, \nu(d\mathbf{x})$$

over E^{b-1} with respect to $d\mathbf{x}_{\hat{h}}$ we obtain, using independence of Z and T,

$$P[Z_1 \in dx_h] = P[Z_{T(h)} \in dx_h \mid T(h) = 1] = g(x_h)\, \nu_h(dx_h),$$

where ν_h is the projection of ν onto the hth factor of E^b. It follows that ν_h is σ-finite on the set $F = \{x \in E / g(x) > 0\}$ and that g is strictly positive P_{Z_1}-a.s.. Now, the assumption implies

$$\begin{aligned} &g(x_1)\, P[Z_T \in d\mathbf{x} \mid T(h) = 1] \\ &\quad = g(x_1)\, g(x_h)\, \nu(d\mathbf{x}) = g(x_h)\, P[Z_T \in d\mathbf{x} \mid T(1) = 1] \end{aligned}$$

which is (ii) with $f = g$.

It remains to prove that (γ) implies (α). Putting $\rho = P_{Z_1} / f$, we show that $D^\rho(\mathbf{x}, h) = D^\rho(\mathbf{x}, k)$ for μ-a.a $\mathbf{x} \in E_h \cap E_k$ which will imply Condition ($\mathscr{C}^\rho_{\mathbf{x}}$) for μ-a.a $\mathbf{x} \in E^b$. (Here, as always, μ is as in (5).) By 3.3(b) and (ii), we have

$$\begin{aligned} f(x_1)\, P[Z_1 \in dx_h]\, K_h(x_h, d\mathbf{x}_{\hat{h}}) &= f(x_1)\, P[Z_T \in d\mathbf{x} \mid T(h) = 1] \\ &= f(x_h)\, P[Z_T \in d\mathbf{x} \mid T(1) = 1] \\ &= f(x_h)\, P[Z_1 \in dx_1]\, K_1(x_1, d\mathbf{x}_{\hat{1}}). \end{aligned}$$

Applying this equality to k instead of h and using (ii) again, we obtain

$$\begin{aligned} f(x_h)\, f(x_1)\, P[Z_1 \in dx_k]\, K_k(x_k, d\mathbf{x}_{\hat{k}}) &= f(x_h)\, f(x_k)\, P[Z_1 \in dx_1]\, K_1(x_1, d\mathbf{x}_{\hat{1}}) \\ &= f(x_k)\, f(x_1)\, P[Z_1 \in dx_h]\, K_h(x_h, d\mathbf{x}_{\hat{h}}) \end{aligned}$$

and, thus,

$$f(x_h)\, P[Z_1 \in dx_k]\, K_k(x_k, d\mathbf{x}_{\hat{k}}) = f(x_k)\, P[Z_1 \in dx_h]\, K_h(x_h, d\mathbf{x}_{\hat{h}}) \tag{13}$$

for all h and k. This means that, for $\mathbf{x} \in E_h \cap E_k$, we have

$$\rho(dx_k)\, K_k(x_k, d\mathbf{x}_{\hat{k}}) = \frac{1}{f(x_k)}\, P[Z_1 \in dx_k]\, K_k(x_k, d\mathbf{x}_{\hat{k}})$$

$$= \frac{1}{f(x_h)}\, P[Z_1 \in dx_h]\, K_h(x_h, d\mathbf{x}_{\hat{h}}) = \rho(dx_h)\, K_h(x_h, d\mathbf{x}_{\hat{h}});$$

this is the claim. ▮

The equivalence of Parts (α) and (δ) in Theorem 3.6 indicates how to construct a reference measure ρ from a suitable measure ν on E^b. We propose two different approaches to verify Statement (δ). The first is exchangeability and the second regularity of T as introduced below. A measure on a product space E^n is *exchangeable* if it remains unchanged under any permutation of the n coordinates.

3.7. *Corollary*

Let ν be an exchangeable measure on E^b such that $P_Z \ll \nu$ (e.g. $\nu = \mu$ in (6)). If $P_Z(d\mathbf{x})/\nu(d\mathbf{x}) = g(x_1)$ is a function of x_1 alone then we have $(\mathscr{C}^{\rho}_{\mathbf{x}})$ with $\rho = P_{Z_1}/g$ for μ-a.a. $\mathbf{x} \in E^b$.

Proof. By exchangeability of ν and the hypothesis we have for all $h \in 1..b$

$$\frac{P[Z_T \in d\mathbf{x} \mid T(h) = 1]}{\nu(d\mathbf{x})} = \frac{1}{q_h} \frac{P[Z_T \in d\mathbf{x}, T(h) = 1]}{\nu(d\mathbf{x})} \tag{14}$$

$$= \frac{1}{q_h} \sum_{\pi:\, \pi(h) = 1} \frac{P[Z_\pi \in d\mathbf{x}]}{\nu(d\mathbf{x})}\, P[T = \pi]$$

$$= \frac{1}{q_h} \sum_{\pi:\, \pi(h) = 1} \frac{P[Z \in d\mathbf{x}_{\pi^{-1}}]}{\nu(d\mathbf{x}_{\pi^{-1}})}\, P[T = \pi]$$

$$= \frac{1}{q_h} \sum_{\pi:\, \pi(h) = 1} g(x_h)\, P[T = \pi] = g(x_h).$$

This shows 3.6(δ) and the corollary follows from Theorem 3.6. ▮

The joint random variable Z is *exchangeable conditional on* Z_1 if the distribution $P[Z_{\hat{1}} \in \cdot \mid Z_1 = x]$ on E^{b-1} is exchangeable for P_{Z_1}-a.a. $x \in E$. Since, for any permutation $\pi \in \mathscr{S}_{b,1}$, we have

$$P[Z_\pi \in d\mathbf{x}] = P[(Z_\pi)_{\hat{1}} \in dx_{\hat{1}} \mid Z_1 = x_1]\, P[Z_1 \in dx_1],$$

the characterization of the conditional probability shows that this conditional exchangeability is tantamount to equivalence of Z_π and Z for all such permutations. From the assumptions of Corollary 3.7

it follows $P_{Z_\pi}(d\mathbf{x}) = g(x_1)\, \nu_\pi(d\mathbf{x}) = P_Z(d\mathbf{x})$ for all $\pi \in \mathscr{S}_{b,1}$. Hence Z is exchangeable conditional on Z_1, here.

The last corollary can easily be applied to the following independent case.

3.8. *Corollary*

Let all b variants be independent, let $Z_2, \ldots, Z_b$ be identically distributed, and suppose that P_{Z_1} is absolutely continuous with respect to P_{Z_2}. Then $(\mathscr{C}_{\mathbf{x}}^{\rho})$ is satisfied with $\rho = P_{Z_2}$ for $(P_{Z_2})^{\otimes b}$-a.a. $\mathbf{x} \in E^b$.

Regularity of T. We call the random permutation T *regular* if the probability $P[T=\pi]$ depends on the site $\pi^{-1}(1)$ of the regular variant, only. Since simple selectors use mainly the properties of the regular variant, regularity of T is a natural assumption in the present context. Plainly, T is regular if and only if P_T is uniform on each set $\mathscr{S}_{b,h}$, i.e., if $P[T=\pi] = q_h/(b-1)!$ when $\pi(h)=1$. Of course, if $b=2$ then T is always regular.

Let us adapt 3.6(δ) to the case of a regular T. Let $S\colon (\Omega, P) \to \mathscr{S}_{b,1}$ be a uniformly distributed random permutation, independent of Z. By independence, we have, for all $h \in 1..b$,

$$
\begin{aligned}
(15)\quad & P[Z_T \in d\mathbf{x} \mid T(h) = 1] \\
&= \sum_{\pi \in \mathscr{S}_{b,h}} P[Z_\pi \in d\mathbf{x}]\, \frac{P[T=\pi]}{q_h} = \frac{1}{(b-1)!} \sum_{\pi \in \mathscr{S}_{b,h}} P[Z_\pi \in d\mathbf{x}] \\
&= \sum_{\pi \in \mathscr{S}_{b,1}} P[(Z_\pi)_{(1,h)} \in d\mathbf{x},\, S=\pi] = P[(Z_S)_{(1,h)} \in d\mathbf{x}].
\end{aligned}
$$

Therefore, if T is regular then 3.6(i) assumes the simpler form

$$f(x_h)\, P[(Z_S)_{(1,k)} \in d\mathbf{x}] = f(x_k)\, P[(Z_S)_{(1,h)} \in d\mathbf{x}], \qquad h, k \in 1..b.$$

3.9. *Corollary*

Let T be regular and let ν be an exchangeable measure on E^b such that $P_{Z_S} \ll \nu$ (e.g. $\nu = \mu$ in (6)). If $P_{Z_S}(d\mathbf{x})/\nu(d\mathbf{x}) = g(x_1)$ depends on the first argument x_1 of $\mathbf{x}$ alone then we have $(\mathscr{C}_{\mathbf{x}}^{\rho})$ with $\rho = P_{Z_1}/g$ for μ-a.a. $\mathbf{x} \in E^b$.

Proof. By (15) together with exchangeability of ν and the hypothesis, we have for all $h \in 1..b$

$$\frac{P[Z_T \in d\mathbf{x} \mid T(h)=1]}{\nu(d\mathbf{x})} = \frac{P[(Z_S)_{(1,h)} \in d\mathbf{x}]}{\nu(d\mathbf{x})} = \frac{P[Z_S \in d\mathbf{x}_{(1,h)}]}{\nu(d\mathbf{x}_{(1,h)})} = g(x_h).$$

This shows 3.6(δ) and the corollary follows from Theorem 3.6. ∎

3.10. *Corollary*

Let T be regular and assume that the b random variables $Z_1, Z_{S(2)}, \ldots, Z_{S(b)}$ are independent. If $P_{Z_1} \ll P_{Z_{S(2)}}$ then $(\mathscr{C}_{\mathbf{x}}^{\rho})$ is satisfied with $\rho = P_{Z_{S(2)}}$ for $(P_{Z_{S(2)}})^{\otimes b}$-a.a. $\mathbf{x} \in E^b$.

Proof. We may put $\nu(d\mathbf{x}) := (P_{Z_{S(2)}} \otimes P_{Z_{S(2)}} \otimes \cdots \otimes P_{Z_{S(b)}}) = (P_{Z_{S(2)}})^{\otimes b}$ and the claim follows from Corollary 3.9. ▌

3.11. *Examples*

(a) Let E be finite or a bounded interval in $\mathbb{R}$ and suppose that T is regular. If the regular variant Z_1 is independent of the joint irregular variants $(Z_2, \ldots, Z_b)$, if $P_{Z_1} \ll P_{Z_{S(2)}}$, and if $(Z_{S(2)}, \ldots, Z_{S(b)})$ is uniformly distributed on E^{b-1} then the assumptions of Corollary 3.10 are satisfied. Corollary 3.8 may not be applicable.

(b) There exist random variables $Z_1, \ldots, Z_b$ not of the form (a) such that the assumptions of Corollary 3.10 are satisfied but Corollary 3.8 fails to be applicable. Consider the following example: Let $b = 3$, let $E = 0..1$, and let the regular variant $Z_1 : (\Omega, P) \to E$ be independent of (Z_2, Z_3): $(\Omega, P) \to E \times E$. Furthermore, let the joint distribution of (Z_2, Z_3) be defined by

$$P[Z_2 = 0, Z_3 = 0] = \tfrac{1}{2}, \qquad P[Z_2 = 1, Z_3 = 0] = 0,$$
$$P[Z_2 = 0, Z_3 = 1] = \sqrt{2} - 1, \qquad P[Z_2 = 1, Z_3 = 1] = \frac{3}{2} - \sqrt{2}.$$

The irregular variants Z_2, Z_3 are neither independent nor identically distributed. However, $P_{(Z_{S(2)}, Z_{S(3)})} = \frac{1}{2}[P_{(Z_2, Z_3)} + P_{(Z_3, Z_2)}]$ and simple algebraic manipulations show that this is the tensor square of the distribution $Z_{S(2)} = (\sqrt{2}/2, 1 - \sqrt{2}/2)$.

This concludes our discussion of Theorem 3.6. The following theorem suggests how to construct a reference measure ρ such that $(\mathscr{C}_{\mathbf{x}}^{\rho})$ holds starting from the kernels K_h.

3.12. *Theorem*

Let ρ be a σ-finite measure on E such that $P_{Z_1} \ll \rho$. If the measure $\rho \otimes K_h$ on E^b is exchangeable and independent of h then we have $(\mathscr{C}_{\mathbf{x}}^{\rho})$ for μ-a.a. $\mathbf{x} \in E^b$.

Proof. Using the assumptions on $\rho \otimes K_h$, we infer from 3.1(b)

$$\begin{aligned}\mu(d\mathbf{x})\, D^\rho(\mathbf{x}, h) &= \rho \otimes K_h(x_h, d\mathbf{x}_{\hat{h}}) \\ &= \rho \otimes K_1(x_h, d\mathbf{x}_{\hat{h}}) \\ &= \rho \otimes K_1(x_1, d\mathbf{x}_{\hat{1}}) = \mu(d\mathbf{x})\, D^\rho(\mathbf{x}, 1).\end{aligned}$$

This equality implies $D^\rho(\cdot, h) = D^\rho(\cdot, 1)$ μ-a.s. establishing $(\mathscr{C}_{\mathbf{x}}^\rho)$ for μ-a.a. $\mathbf{x} \in E^b$. ∎

Let us next draw some conclusions from Theorem 3.12. We deal first with conditions guaranteeing existence of a measure ρ so that the product $\rho \otimes K_h$ is exchangeable; we will subsequently deal with the question of equality of the measures $\rho \otimes K_h$, $h \in 1..b$. A Markov kernel N: $E \times \mathscr{B}(E) \to [0, 1]$ is *reversible* with *reversible measure* ρ on $\mathscr{B}(E)$ if the *equation of detailed balance*

$$\rho(dx)\, N(x, dy) = \rho(dy)\, N(y, dx)$$

holds, i.e., if the measure $\rho \otimes N$ is *exchangeable* on $E \times E$. Now, given a kernel M: $E \times \mathscr{B}(E^{b-1}) \to [0, 1]$, let us investigate, under what conditions $\rho \otimes M$ is exchangeable if $b > 2$. To this end, we derive from M two more Markovian kernels. The first, M': $E \times \mathscr{B}(E) \to [0, 1]$, is the projection of M defined by

$$M'(x, dy) = M(x, dy \times E^{b-1}), \qquad x \in E.$$

Since E is Polish, there is also a Markovian kernel M_ρ: $(E \times E) \times \mathscr{B}(E^{b-2}) \to [0, 1]$ associated with ρ and M such that

$$M_\rho((x, y), d\mathbf{z}) = (\rho \otimes M)[(\iota_3, ..., \iota_b) \in d\mathbf{z} \mid \iota_1 = x, \iota_2 = y];$$

here, ι_k is the projection of E^b onto the kth factor. By the product formula for conditional probabilities, we have

$$M(x, dy\, d\mathbf{z}) = M'(x, dy)\, M_\rho((x, y), d\mathbf{z}) \tag{16}$$

for ρ-a.a. $x \in E$. Note that M_ρ is essentially a function of M, namely

$$M_\rho((x, y), d\mathbf{z}) = M(x, \cdot)[(\iota_3, ..., \iota_b) \in d\mathbf{z} \mid \iota_2 = y].$$

(Had we defined M_ρ by this formula, simultaneous measurability of M_ρ in x and y would not be guaranteed.) It is these kernels that appear in the following lemma.

3.13. *Lemma*

If $b > 2$ then the following are equivalent for a σ-finite measure ρ on E.

(a) The measure $\rho \otimes M$ is exchangeable.
(b) (i) $M(x, \cdot)$ is exchangeable for ρ-a.a. $x \in E$;
(ii) the measure $\rho \otimes M$ is symmetric in its first two coordinates.
(c) (i) $M(x, \cdot)$ is exchangeable for ρ-a.a. $x \in E$;
(iii) ρ is reversible with respect to M';
(iv) $M_\rho((x, y), \cdot)$ is symmetric in x and y for $\rho(dx)\, M'(x, dy)$-a.a. $(x, y) \in E \times E$.

Proof. Part (i) of the implication (a) $\Rightarrow$ (b) follows from conditioning $\rho \otimes M$ on the first coordinate and (ii) is immediate. Assume now (b). Integrating the equality $\rho(dx_1)\, M(x_1, d\mathbf{x}_{\hat{1}}) = \rho(dx_2)\, M(x_2, d\mathbf{x}_{\hat{2}})$ over the last $b-2$ coordinates we obtain the equation of detailed balance for ρ and M'; i.e., (iii). In order to prove (iv) we use symmetry (ii) and (16) to compute

$$
\begin{aligned}
&\rho(dx_2)\, M'(x_2, dx_1)\, M_\rho((x_2, x_1), d\mathbf{x}_{\widehat{12}})\\
&\quad= \rho(dx_2)\, M(x_2, d\mathbf{x}_{\hat{2}})\\
&\quad= \rho(dx_1)\, M(x_1, d\mathbf{x}_{\hat{1}})\\
&\quad= \rho(dx_1)\, M'(x_1, dx_2)\, M_\rho((x_1, x_2), d\mathbf{x}_{\widehat{12}}).
\end{aligned}
$$

Part (c) now follows from (iii).

Turning to the implication (c) $\Rightarrow$ (a) we let now ρ be as in (c). By (i), it is sufficient to prove that $\rho(dx_h)\, M(x_h, \mathbf{x}_{\hat{h}})$ does not depend on $h \in 1..b$. This follows from the subsequent chain of equations which uses (16), (iv), (iii), and (i) in this order.

$$
\begin{aligned}
&\rho(dx_h)\, M(x_h, d\mathbf{x}_{\hat{h}})\\
&\quad= \rho(dx_h)\, M'(x_h, dx_1)\, M_\rho((x_h, x_1), d\mathbf{x}_{\widehat{1h}})\\
&\quad= \rho(dx_h)\, M'(x_h, dx_1)\, M_\rho((x_1, x_h), d\mathbf{x}_{\widehat{1h}})\\
&\quad= \rho(dx_1)\, M'(x_1, dx_h)\, M_\rho((x_1, x_h), d\mathbf{x}_{\widehat{1h}})\\
&\quad= \rho(dx_1)\, M(x_1, dx_h\, d\mathbf{x}_{\widehat{1h}})\\
&\quad= \rho(dx_1)\, M(x_1, d\mathbf{x}_{\hat{1}}). \quad \blacksquare
\end{aligned}
$$

We next propose two different approaches to equality of the measures $\rho \otimes K_h$, $h \in 1..b$, appearing in Theorem 3.12. The first one is exchangeability of $K(x, \cdot)$ for ρ-a.a. $x \in E$, cf. 3.13(i), and the second one regularity of T. From the former condition and (7), we first deduce

(17) $$K_h(x, \cdot) = K(x, \cdot), \qquad h \in 1..b, \ \rho\text{-a.a. } \mathbf{x}.$$

The next corollary flows directly from this equality and Theorem 3.12.

3.14. *Corollary*

Let ρ be a σ-finite measure on E such that $P_{Z_1} \ll \rho$.

(a) If $\rho \otimes K$ is exchangeable then we have $(\mathscr{C}_{\mathbf{x}}^{\rho})$ for μ-a.a. $\mathbf{x} \in E^b$.

(b) If $b = 2$ and if the Markovian kernel K is reversible with reversible measure ρ then we have $(\mathscr{C}_{\mathbf{x}}^{\rho})$ for μ-a.a. $\mathbf{x} \in E^2$. ∎

Let us define the symmetrized kernel $L: E \times \mathscr{B}(E^{b-1}) \to [0, 1]$ of K,

(18) $$L(x, d\mathbf{y}) := \frac{1}{(b-1)!} \sum_{\sigma \in \mathscr{S}_{b-1}} K(x, d\mathbf{y}_\sigma).$$

If T is regular as already defined above then the restriction of P_T on $\mathscr{S}_{b,h}$ is uniform; hence, we obtain from Formula (7)

(19) $$K_h(x, \cdot) = L(x, \cdot), \qquad h \in 1..b, \ \rho\text{-a.a. } x \in E.$$

If $b = 2$ then, plainly, $K_h = K = L$.

Our next corollaries follow from Theorem 3.12 and (19).

3.15. *Corollary*

Let T be regular and let ρ be a σ-finite measure on E such that $P_{Z_1} \ll \rho$ and $\rho \otimes L$ is exchangeable. Then we have $(\mathscr{C}_{\mathbf{x}}^{\rho})$ for μ-a.a. $\mathbf{x} \in E^b$.

Another interesting situation related to Theorem 3.12 arises when the irregular variants are iterated functions of the regular one. Let $Z_{j+1} = \varphi^j(Z_1)$, $j < b$, for some function $\varphi: E \to E$. The elements in E^{b-1} *observable* for X are then of the form $(\varphi^{\pi(1)-1}(x), ..., \varphi^{\pi(b)-1}(x))$ with some permutation $\pi \in \mathscr{S}_b$ and some $x \in E$. If, in addition, φ^b is the identity function on E then they are all of the form $(y, \varphi^{\sigma(1)}(y), ..., \varphi^{\sigma(b-1)}(y))$ with some permutation $\sigma \in \mathscr{S}_{b-1}$ and $y = \varphi^{\pi(1)-1}(x)$. If $b = 2$ then φ is an involution.

3.16. *Corollary*

Assume that T is regular. Let $Z_{j+1} = \varphi^j(Z_1)$, $j < b$, for some measurable function $\varphi: E \to E$ such that φ^b is the identity function, and let ρ be a

φ-invariant reference measure such that $P_{Z_1} \ll \rho$. Then, we have $(\mathscr{C}_{\mathbf{x}}^{\rho})$ for all observable $\mathbf{x} \in E^b$ such that x_1 is off some ρ-nullset.

Proof. We use the kernel

$$K(x_1, d(x_2, \ldots, x_b)) = \prod_{j=1}^{b-1} \delta_{\varphi(x_j)}(x_{j+1}) = \prod_{j=1}^{b-1} \delta_{\varphi^j(x_1)}(x_{j+1}) \tag{20}$$

and its symmetrization (18)

$$\begin{aligned} L(x_1, d(x_2, \ldots, x_b)) &= \frac{1}{(b-1)!} \sum_{\sigma \in \mathscr{S}_{b,1}} K(x_1, d(x_{\sigma(2)}, \ldots, x_{\sigma(b)})) \\ &= \frac{1}{(b-1)!} \sum_{\sigma \in \mathscr{S}_{b,1}} \prod_{j=1}^{b-1} \delta_{\varphi^j(x_1)}(x_{\sigma(j+1)}). \end{aligned} \tag{21}$$

Invoking Corollary 3.15, we establish 3.13(ii) with $M = L$, i.e. symmetry of $\rho \otimes L$ in the first two coordinates (recall that $L(x, \cdot)$ is exchangeable). Indeed, for any bounded, measurable function $f: E^b \to \mathbb{R}$, (21) first implies

$$\begin{aligned} &\int \rho(dx_1) \int L(x_1, d(x_2, \ldots, x_b))\, f(\mathbf{x}) \\ &= \frac{1}{(b-1)!} \sum_{\sigma \in \mathscr{S}_{b-1}} \int \rho(dx_1)\, f(x_1, \varphi^{\sigma(1)}(x_1), \varphi^{\sigma(2)}(x_1), \ldots, \varphi^{\sigma(b-1)}(x_1)). \end{aligned} \tag{22}$$

Now, as σ runs through all permutations of $1..(b-1)$, so does the permutation

$$\eta = \begin{cases} 1 \mapsto -\sigma(1) \mod b, & \\ j \mapsto \sigma(j) - \sigma(1) \mod b, & 2 \leqslant j < b, \end{cases}$$

Thus, using φ-invariance of ρ and $\varphi^b =$ id, we obtain

$$\begin{aligned} (22) &= \frac{1}{(b-1)!} \sum_{\sigma \in \mathscr{S}_{b-1}} \int \rho(dx_1)\, f(\varphi^{-\sigma(1)}(x_1), x_1, \varphi^{\sigma(2)-\sigma(1)}(x_1), \ldots, \varphi^{\sigma(b-1)-\sigma(1)}(x_1)) \\ &= \frac{1}{(b-1)!} \sum_{\eta \in \mathscr{S}_{b-1}} \int \rho(dx_1)\, f(\varphi^{\eta(1)}(x_1), x_1, \varphi^{\eta(2)}(x_1), \ldots, \varphi^{\eta(b-1)}(x_1)). \end{aligned} \tag{23}$$

A comparison of (22) and (23) proves the claim. ∎

The counting measure # on E is φ-invariant for any bijective φ. Therefore, there is also the following corollary.

3.17. *Corollary*

Let E be countable. If T is regular and if $Z_{j+1} = \varphi^j(Z_1)$, $j < b$, for some $\varphi: E \to E$ such that φ^b is the identity function then we have $(\mathscr{C}_{\mathbf{x}}^{\#})$ for all observable $\mathbf{x} \in E^b$. ∎

3.18. *Remarks*

(a) Theorem 3.6 and Theorem 3.12 and their corollaries offer ways of establishing Condition $(\mathscr{C}_{\mathbf{x}}^{\rho})$. Whereas 3.6($\delta$) resorts to the measure ν in order to construct a suitable reference measure ρ, Theorem 3.12 needs no such input. It rather indicates how to construct ρ from the distribution P_{Z_T}. Note also that the kernels K_h, K, and L and, hence, the reference measures ρ appearing in Theorem 3.12 and its corollaries depend on Z_1 through the requirement $P_{Z_1} \ll \rho$, only.

(b) It follows from Theorem 3.4(b) that the hypotheses of Theorems 3.6 and 3.12 and of their corollaries all imply the equality $SS^{\rho} = BS$, P_{Z_T}-a.s..

(c) The function f in Theorem 3.6 and the measures ρ in Corollary 3.14(a),(b) are related. It is an easy consequence of Lemma 3.3(b) that $f := dP_{Z_1}/d\rho$ satisfies 3.6(i). The converse is not true. If f satisfies 3.6(i) then the measure $\frac{P_{Z_1}}{f} \otimes K$ is not necessarily exchangeable. The support of P_{Z_1} may be too small. It is sufficient to consider $E = 0..1$, $b = 2$, $P_Z = \frac{1}{2}(\delta_{(0,0)} + \delta_{(0,1)})$, $f = 1_{\{0\}}$, and $K(x, \cdot) = (\frac{1}{2}, \frac{1}{2})$ for $x = 0, 1$. The measure $\frac{P_{Z_1}}{f} = \delta_0$ violates the equation of detailed balance.

4. IRREGULAR VARIANTS FROM GAUSSIAN NOISE

In this section, we apply Corollary 3.14 and Lemma 3.13(c) to the Gaussian case. More precisely, we let $E = \mathbf{R}^d$ and assume that the irregular variants arise from the regular variant as affine transformations corrupted by nondegenerate Gaussian noise, i.e., from the Gaussian linear model

$$\begin{pmatrix} Z_2 \\ \vdots \\ Z_b \end{pmatrix} = \mathbf{m} + \mathbf{S} Z_1 + R. \tag{24}$$

Here, $\mathbf{m} \in \mathbf{R}^{(b-1)\,d}$, $\mathbf{S} \in \mathbf{R}^{(b-1)\,d \times d}$, $R \sim N(0, \mathbf{V})$ is independent of Z_1, and $\mathbf{V}$ is a positive definite $(b-1)\,d$ by $(b-1)\,d$ matrix. More generally, we assume

$$K(x, \cdot) \sim N(\mathbf{m} + \mathbf{S}x, \mathbf{V}), \qquad x \in \mathbf{R}^d. \tag{25}$$

4.1. *Lemma*

Let $K'\colon \mathbf{R}^d \times \mathscr{B}(\mathbf{R}^d) \to [0, 1]$ be the Gaussian transition kernel defined by $K'(x, \cdot) \sim N(m + Sx, V)$, $m \in \mathbf{R}^d$, $S, V \in \mathbf{R}^{d \times d}$, V regular. K' is reversible if and only if SV is symmetric. In this case, the reversible measure is unique up to a factor and has Lebesgue density

$$e^{-\frac{1}{2}\{(x-m)^T V^{-1}(x-m) - (m+Sx)^T V^{-1}(m+Sx)\}}. \tag{26}$$

Proof. A reversible measure of K' has an everywhere strictly positive Lebesgue density. The lemma now follows from a straightforward computation. ∎

The following remark is obvious.

4.2. *Remark*

A transition kernel $K(x, \cdot)$ of the form (25) is exchangeable for all $x \in \mathbf{R}^d$ if and only if $\mathbf{m}$, $\mathbf{S}$, and $\mathbf{V}$ are of the form

$$\mathbf{m} = \begin{pmatrix} m \\ \vdots \\ m \end{pmatrix}, \qquad m \in \mathbf{R}^d, \tag{27}$$

$$\mathbf{S} = \begin{pmatrix} S \\ \vdots \\ S \end{pmatrix}, \qquad S \in \mathbf{R}^{d \times d}, \tag{28}$$

$$\mathbf{V} = \begin{pmatrix} V & U & \cdots & U \\ U & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & U \\ U & \cdots & U & V \end{pmatrix}, \tag{29}$$

where $U \in \mathbf{R}^{d \times d}$ symmetric (i.e., $\mathbf{V}$ is block circulant). We will need the following lemma on block-circulant matrices.

4.3. *Lemma*

Let $r \geqslant 2$ be a natural number, let V and U be two d by d matrices and let $\mathbf{V}$ be an rd by rd block-circulant matrix

$$\mathbf{V} = \begin{pmatrix} V & U & \cdots & U \\ U & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & U \\ U & \cdots & U & V \end{pmatrix}.$$

(a) The determinant of $\mathbf{V}$ is given by

$$\det \mathbf{V} = \det[V-U]^{r-1} \det[V+(r-1)\,U]. \tag{30}$$

(b) $\mathbf{V}$ is regular if and only if both matrices $V-U$ and $V+(r-1)\,U$ are regular and then the inverse of $\mathbf{V}$ is

$$\mathbf{V}^{-1} = \begin{pmatrix} B & C & \cdots & C \\ C & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & C \\ C & \cdots & C & B \end{pmatrix}, \tag{31}$$

where

$$B = \frac{1}{r}\left([V+(r-1)\,U]^{-1} + (r-1)[V-U]^{-1}\right), \tag{32}$$

$$C = \frac{1}{r}\left([V+(r-1)\,U]^{-1} - [V-U]^{-1}\right). \tag{33}$$

Moreover, we have

$$B+(r-1)\,C = [V+(r-1)\,U]^{-1}, \tag{34}$$

$$B-C = [V-U]^{-1}. \tag{35}$$

Letting $G = -(B+(r-1)\,C)^{-1}\,C$, we have also

$$U = G[V-U], \tag{36}$$

$$G+I = V[B-C], \tag{37}$$

$$V^{-1} - G^T V^{-1} G = B + rCG. \tag{38}$$

(c) Equalities (32)–(35) subsist after swapping V for B and U for C.

Proof. The representation of the determinant $\det \mathbf{V}$ appears in [3], Theorem 8.9.1, that of $\mathbf{V}^{-1}$ with the Equalities (32) and (33) is verified by a direct computation. Equalities (34) and (35) follow from (32) and (33).

According to (33), (34), and the definition of G, we have

$$[V+(r-1)\,U]^{-1}\,G=\frac{1}{r}\,([V-U]^{-1}-[V+(r-1)\,U]^{-1}).$$

This can be rewritten as

$$V+(r-1)\,U=(I+rG)[V-U],$$

i.e., (36).

Equation (37) follows directly from the identity $I=(V-U)(B-C)$ and (36). In order to obtain (38), use the formula $G^TV^{-1}=B-C-V^{-1}$, cf. (37), and the definition of G to compute

$$\begin{aligned}(39)\qquad I-VG^TV^{-1}G-VB-rVCG&=I-V(B-C-V^{-1})\,G-VB-rVCG\\&=I-V(B+(r-1)\,C)\,G+G-VB\\&=I+G-V[B-C]\\&=0,\end{aligned}$$

again by (37). The claim follows from regularity of V. ∎

4.4. *Proposition*

Let K be of the form (25).

(a) There exists a σ-finite measure ρ on $\mathbf{R}^d$ such that $\rho\otimes K$ is exchangeable if and only if $\mathbf{m}$, $\mathbf{S}$, and $\mathbf{V}$ are of the form (27), (28), (29), respectively, and if

$$(40)\qquad \begin{cases} SV \text{ is symmetric}, & \text{if } \ b=2,\\ U=S[V-U], & \text{if } \ b\geqslant 3.\end{cases}$$

In this case, ρ is unique up to a factor and has Lebesgue density (26).

(b) If Z_1 possesses a Lebesgue density and if (27), (28), (29), and (40) are satisfied then we have $(\mathscr{C}^{\rho}_{\mathbf{x}})$ for Lebesgue a.a. $\mathbf{x}\in\mathbf{R}^{bd}$.

Proof. Let us first note that (40) implies symmetry of SV also in the case $b\geqslant 3$. Indeed, in this case, Condition (40) is equivalent to $(S+I)(V-U)=V$ and, hence, to $V^{-1}S=(V-U)^{-1}-V^{-1}$. Since both matrices V and U are symmetric, so is $V^{-1}S$ and, hence, also SV.

We apply Corollary 3.14 in order to prove (a). Let us verify 3.13(c) with $M = K$. We have already remarked that 3.13(i) is equivalent to the validity of (27), (28), and (29). Now, $K'(x, dy) = P[Z_2 \in dy \mid Z_1 = x] = N(m + Sx, V)(dy)$. Hence, by Lemma 4.1, Condition 3.13(iii) is equivalent to symmetry of SV; uniqueness and the representation (26) of ρ follow as well. We have thus proved Part (a) in the case $b = 2$ and it remains to show that 3.13(iv) is equivalent to (40) if $b \geqslant 3$ and if (27), (28), and (29) are valid. Let us first show that (40) is *necessary* for having 3.13(iv). We have

$$\rho(dx)\, K(x, dy\, d\mathbf{z}) = e^{-\frac{1}{2} r(x)} e^{-\frac{1}{2}\{((^{y}_{\mathbf{z}}) - \mathbf{m} - \mathbf{S}x)^T\, \mathbf{V}^{-1}((^{y}_{\mathbf{z}}) - \mathbf{m} - \mathbf{S}x)\}}\, dx\, dy\, d\mathbf{z},$$

where $r(x)$ is the expression in curly brackets appearing in (26). Hence, 3.13(iv) says that the expression

$$r(x) + \left(\begin{pmatrix} y \\ \mathbf{z} \end{pmatrix} - \mathbf{m} - \mathbf{S}x\right)^T \mathbf{V}^{-1} \left(\begin{pmatrix} y \\ \mathbf{z} \end{pmatrix} - \mathbf{m} - \mathbf{S}x\right) \tag{41}$$

is symmetric in x and y for all $\mathbf{z} \in \mathbf{R}^{(b-2)\,d}$. The terms containing x, z_i and y, z_i in this polynomial are $-2\,(0, \mathbf{z}^T)\,\mathbf{V}^{-1}\mathbf{S}x$ and $2\,(0, \mathbf{z}^T)\,\mathbf{V}^{-1}(^{y}_{\mathbf{0}})$, respectively. Now,

$$(0, \mathbf{z}^T)\,\mathbf{V}^{-1} = \mathbf{z}^T \begin{pmatrix} C & B & C & \cdots & C \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & C \\ C & \cdots & \cdots & C & B \end{pmatrix},$$

cf. (31). A necessary condition for the symmetry of this polynomial in x and y is, therefore,

$$(B + (b-2)\, C)\, S = -C, \tag{42}$$

i.e., that S be the matrix G introduced in Lemma 4.3. Equation (36) now implies (40).

We finally show that the conditions (27), (28), (29), and (40) are also *sufficient* for symmetry of (41) in x and y. Let us start with Equation (42) instead of (40); it is equivalent to

$$\mathbf{V}^{-1}\mathbf{S} = -\mathbf{C} := -(C, \ldots, C)^T \in \mathbf{R}^{(b-1)\,d \times d},$$

an equality that we use in order to reformulate (41) as

(43)

$$r(x)+\left(\begin{pmatrix} y\\ \mathbf{z}\end{pmatrix}-\mathbf{m}\right)^T \mathbf{V}^{-1}\left(\begin{pmatrix} y\\ \mathbf{z}\end{pmatrix}-\mathbf{m}\right)-(b-1)\, x^T CSx+2x^T\mathbf{C}^T\left(\begin{pmatrix} y\\ \mathbf{z}\end{pmatrix}-\mathbf{m}\right).$$

The term containing x and y in this polynomial is $-2x^TCy$; it is symmetric since C is. The term containing the powers of x is

$$x^TV^{-1}x-2m^TV^{-1}x-x^TS^TV^{-1}Sx-2m^TV^{-1}Sx-(b-1)\, x^TCSx-2(b-1)\, m^TCx$$

and the one containing the powers of y is

$$y^TBy-2m^T(B+(b-2)\, C)\, y.$$

Symmetry of these terms finally follows from (38) and (37). Claim (b) is now a consequence of Corollary 3.14. ▮

4.5. *Corollary*

Let $Z=(Z_1, ..., Z_b)\sim N(\mathbf{m}, \mathbf{W})$ with $\mathbf{m}=(m_1, m_2, ..., m_2)\in\mathbf{R}^{bd}$ and a nonsingular covariance matrix of the form

(44)
$$\mathbf{W}=\begin{pmatrix} W & H^T & \cdots & \cdots & H^T\\ H & M & L & \cdots & L\\ \vdots & L & \ddots & \ddots & \vdots\\ \vdots & \vdots & \ddots & \ddots & L\\ H & L & \cdots & L & M\end{pmatrix}=\begin{pmatrix} W & \mathbf{H}^{\mathbf{T}}\\ \mathbf{H} & \mathbf{M}\end{pmatrix}\in\mathbf{R}^{bd\times bd}$$

(in particular, L is symmetric). If

(45)
$$\begin{cases} HW^{-1}(M-HW^{-1}H^T) \text{ is symmetric}, & \text{if } \ b=2,\\ HW^{-1}(M+H^T)=(I+HW^{-1})\, L, & \text{if } \ b>2,\end{cases}$$

then we have $SS^\rho(\mathbf{x})=BS(\mathbf{x})$ for Lebesgue-a.a. $\mathbf{x}\in\mathbf{R}^{bd}$ and some Lebesgue absolutely continuous measure ρ on $\mathbf{R}^d$. Moreover, with the notations

(46) $$A=(W-\mathbf{H}^T\mathbf{M}^{-1}\mathbf{H})^{-1},$$

(47) $$B=\frac{1}{b-1}([M+(b-2)\, L-(b-1)\, HW^{-1}H^T]^{-1}+(b-2)[M-L]^{-1}),$$

(48) $$C=\frac{1}{b-1}([M+(b-2)\, L-(b-1)\, HW^{-1}H^T]^{-1}-[M-L]^{-1}),$$

we have

$$f^{\rho}_{Z_1}(x) = e^{-\frac{1}{2}x^T(A-B)\,x + m_1^T(A-C)\,x + m_2^T(C-B)\,x}. \tag{49}$$

Proof. It follows from [4], Theorem 3.2.4, that

$$P[Z_{\hat{1}} \in \cdot \mid Z_1 = x] = N(\mathbf{m} + \mathbf{S}x, \mathbf{V}) = K(x, \cdot),$$

where $\mathbf{S}$ is of the form (28) with $S = HW^{-1}$, $\mathbf{V} = \mathbf{M} - \mathbf{H}W^{-1}\mathbf{H}^T$ is block circulant of the form (29) with diagonal blocks $V = M - HW^{-1}H^T$ and off-diagonal blocks $U = L - HW^{-1}H^T$ since L is symmetric, and where $\mathbf{m}$ is of the form (27) with $m = m_2 - Sm_1$. Now, Condition (40) follows immediately from (45) and Proposition 4.4 applies.

Let ρ be the measure on $\mathbf{R}^d$ with Lebesgue density (26). It remains to show that

$$f^{\rho}_{Z_1}(x) = e^{-\frac{1}{2}\{(x-m_1)^T W^{-1}(x-m_1) - (x-m)^T V^{-1}(x-m) + (m+Sx)^T V^{-1}(m+Sx)\}} \tag{50}$$

has the representation (49). To this end we need a few preliminaries. Lemma 4.3 shows that $\mathbf{V}^{-1}$ is of the form (31) with B and C given by (47) and (48), cf. (32) and (33). It also follows that S is the matrix G introduced in Lemma 4.3, i.e., we have (42). A simple computation shows the well-known fact that the inverse of W is of the form

$$\mathbf{W}^{-1} = \begin{pmatrix} A & \mathbf{D}^T \\ \mathbf{D} & \mathbf{V}^{-1} \end{pmatrix}, \tag{51}$$

with A and $\mathbf{V}$ as above and

$$\mathbf{D} = -\mathbf{V}^{-1}\mathbf{H}W^{-1} = -\mathbf{V}^{-1}\mathbf{S} = \mathbf{C} \in \mathbf{R}^{(b-1)\,d \times d}, \tag{52}$$

cf. (42).

Now, the matrix A has also the representation

$$A = W^{-1} + W^{-1}\mathbf{H}^T\mathbf{V}^{-1}\mathbf{H}W^{-1} = W^{-1} - \mathbf{D}^T\mathbf{S} = W^{-1} - (b-1)\,CS$$

by (52) and (42). Hence, by (38),

$$W^{-1} - V^{-1} + S^T V^{-1} S = A + (b-1)\,CS - B - (b-1)\,CS = A - B. \tag{53}$$

We already know from (37) that $V^{-1}(I+S) = B - C$. Adding this to (53), we obtain finally

$$\begin{aligned} A-C &= W^{-1}-V^{-1}+S^TV^{-1}S+V^{-1}+V^{-1}S \\ &= W^{-1}+(I+S^T)\,V^{-1}S \\ &= W^{-1}+S^TV^{-1}(I+S) \end{aligned}$$

since SV is symmetric. These equalities, together with $HW^{-1}=S$, show that the expression in curly brackets in (50) equals, up to an additive constant,

$$\begin{aligned} &x^T(W^{-1}-V^{-1}+S^TV^{-1}S)\,x-2m_1^TW^{-1}x+2m^TV^{-1}(I+S)\,x \\ &\quad = x^T(W^{-1}-V^{-1}+S^TV^{-1}S)\,x \\ &\qquad -2m_1^T(W^{-1}+S^TV^{-1}(I+S))\,x+2m_2^TV^{-1}(I+S)\,x \\ &\quad = x^T(A-B)\,x-2m_1^T(A-C)\,x-2m_2^T(C-B)\,x. \end{aligned}$$

This is Claim (49). ∎

4.6. *Corollary*

If all hypotheses of Corollary 4.5 are fulfilled and if $m_1=m_2$ then there exists a Lebesgue absolutely continuous reference measure ρ on $\mathbf{R}^d$ such that $SS^\rho(\mathbf{x})=BS(\mathbf{x})$ for Lebesgue-a.a. $\mathbf{x}\in\mathbf{R}^{bd}$. With respect to this ρ we have

(54) $$f^{\rho}_{Z_1}(x)=e^{-\frac{1}{2}(x-m_1)^T(A-B)(x-m_1)}.$$

4.7. *Discussion*

(a) In the univariate case, Condition (40) is satisfied throughout if $b=2$. If H is symmetric then Condition (45), $b=2$, is satisfied if and only if the matrix $HW^{-1}M$ is symmetric. This is guaranteed, e.g., if H, W, and M commute pairwise or if two of these matrices are equal. If $W+H$ (equivalently $I+HW^{-1}$) is nonsingular then (45), $b>2$, is equivalent to

(55) $$L=H(W+H)^{-1}\,(M+H^T).$$

This is restrictive but not unreasonable. Note also that, since there are only finitely many variants, selectors enjoy robustness to small changes of the model. We do not go into details, here.

(b) Note that the "obvious" simple selector in the situations of Corollary 4.5 or Corollary 4.6,

$$SS^\lambda(\mathbf{x})=\operatorname*{argmax}_h\ q_he^{-\frac{1}{2}(x_h-m_1)^TW^{-1}(x_h-m_1)},$$

is not optimal for all $\mathbf{x}$, cf. Example 2.4.1(b). It corresponds to the density function $f^{\lambda}_{Z_1}$ with Lebesgue measure λ as the reference measure. In Corollary 4.6, the matrix W^{-1} needs a correction term $V^{-1}-S^TV^{-1}S$ as Formula (53) reveals! On the other hand, the conditions given in the Gaussian case are far from being necessary as Example 2.4.1(c) shows.

(c) A *simple* selector needs, besides the distribution P_{Z_1} of the regular variant, the reference measure ρ. In general, there does not exist a reference measure ρ such that SS^{ρ} equals the Bayesian selector, cf. Example 2.4.2. However, the statements in this and the previous section expose interesting situations where such ρ does exist. In these cases, variant selection can be optimally performed by means of the density $f^{\rho}_{Z_1}$ which needs only partial knowledge on the joint distribution of all variants, cf. Theorems 3.6 and 3.12 and their corollaries, and 4.4–4.6. In the case of Proposition 4.5 it needs the symmetric matrix $A-B$ and the vectors $m_1^T(A-C)$ and $m_2^T(C-B)$ instead of the matrices A, B, C and the vectors m_1 and m_2 and Corollary 4.6 just needs $A-B$. In the case of Example 3.5.1 the qualitative information of disjointness is sufficient. In Corollary 3.8 the density function of P_{Z_1} with respect to P_{Z_2} is required and Corollary 3.17 just needs the distribution of the regular variant. Often there is also some freedom as to the choice of ρ, cf. Example 2.4.1(c). Another advantage of simple selectors is that they reduce selection to function evaluation on E instead of E^b.

It is interesting to ask what the assumption (45) appearing in Corollary 4.5 means in terms of the random variables Z_i. Here is an answer.

4.8. *Proposition*

Let $Z=(Z_1, ..., Z_b) \sim N(\mathbf{m}, \mathbf{W})$ with $\mathbf{m} \in \mathbf{R}^{bd}$ and a nonsingular covariance matrix $\mathbf{W}$. Assume that

(i) for all $\pi \in \mathscr{S}_{b,1}$, we have $Z_\pi \sim Z$, i.e., Z is exchangeable conditional on Z_1;

(ii)

$$\begin{cases} \text{the random variables } \operatorname{Var}[Z_2 \mid Z_1=0](\operatorname{Var} Z_1)^{-1} Z_1 & \\ \qquad \text{and } Z_2 \text{ are exchangeable,} & \text{if} \quad b=2, \\ \operatorname{Cov}(\operatorname{Var}[Z_2-Z_3 \mid Z_1=0](\operatorname{Var} Z_1)^{-1} Z_1, Z_2) & \\ \qquad = 2 \operatorname{Cov}[Z_2, Z_3 \mid Z_1=0], & \text{if} \quad b>2. \end{cases}$$

Then the hypotheses of Corollary 4.5 are satisfied.

Proof. It follows from (i) that

$$\mathbf{m} = \begin{pmatrix} EZ_1 \\ EZ_2 \\ \vdots \\ EZ_2 \end{pmatrix}$$

and that **W** is of the form (44) with

(56) $$W = \text{Var } Z_1, \qquad M = \text{Var } Z_2, \qquad H = \text{Cov}(Z_2, Z_1),$$
$$\text{and } L^T = L = \text{Cov}(Z_2, Z_j), \qquad \text{if} \quad b \geqslant j \geqslant 3.$$

It is well known (cf. [4], Theorem 3.2.4) that, for $b \geqslant j \geqslant 3$,

$$\begin{aligned}
\text{Var}&\left[\begin{pmatrix} Z_2 \\ Z_j \end{pmatrix} \Big/ Z_1 = 0\right] \\
&= \text{Var}\begin{pmatrix} Z_2 \\ Z_j \end{pmatrix} - \text{Cov}\left(Z_1, \begin{pmatrix} Z_2 \\ Z_j \end{pmatrix}\right)^T (\text{Var } Z_1)^{-1} \text{Cov}\left(Z_1, \begin{pmatrix} Z_2 \\ Z_j \end{pmatrix}\right) \\
&= \begin{pmatrix} M & L \\ L & M \end{pmatrix} - \begin{pmatrix} H \\ H \end{pmatrix} W^{-1}(H^T H^T) \\
&= \begin{pmatrix} M - HW^{-1}H^T & L - HW^{-1}H^T \\ L - HW^{-1}H^T & M - HW^{-1}H^T \end{pmatrix};
\end{aligned}$$

hence,

(57) $$\text{Var}[Z_i \mid Z_1 = 0] = M - HW^{-1}H^T, \qquad i \geqslant 2,$$

(58) $$\text{Cov}[Z_2, Z_j \mid Z_1 = 0] = L - HW^{-1}H^T, \qquad j > 2.$$

If $b > 2$ then, (57) and (58) imply

(59) $$\begin{aligned}
\text{Var}&[Z_2 - Z_3 \mid Z_1 = 0] \\
&= \text{Var}[Z_2 / Z_1 = 0] + \text{Var}[Z_3 \mid Z_1 = 0] - \text{Cov}[Z_2, Z_3 \mid Z_1 = 0] \\
&\quad - \text{Cov}[Z_3, Z_2 \mid Z_1 = 0] \\
&= 2(M - HW^{-1}H^T) - 2(L - HW^{-1}H^T) \\
&= 2(M - L).
\end{aligned}$$

This equality, (56), (ii), and (58) combine to show

$$\begin{aligned}(M-L)\,W^{-1}H^T &= \tfrac{1}{2}\,\mathrm{Var}[Z_2-Z_3 \mid Z_1=0](\mathrm{Var}\,Z_1)^{-1}\,\mathrm{Cov}(Z_1,Z_2)\\ &= \mathrm{Cov}[Z_2,Z_3 \mid Z_1=0]\\ &= L-HW^{-1}H^T,\end{aligned}$$

i.e., $(M+H)\,W^{-1}H^T = L(I+W^{-1}H^T)$. This is (45) in the case $b>2$.
If $b=2$ then, from (57) and (56), we infer

$$\begin{aligned}(M-HW^{-1}H^T)\,W^{-1}H^T &= \mathrm{Var}[Z_2 \mid Z_1=0](\mathrm{Var}\,Z_1)^{-1}\,\mathrm{Cov}(Z_1,Z_2)\\ &= \mathrm{Cov}(\mathrm{Var}[Z_2 \mid Z_1=0](\mathrm{Var}\,Z_1)^{-1}\,Z_1,Z_2).\end{aligned}$$

Assumption (ii) shows the symmetry of this matrix, i.e. (45) in this case.

ACKNOWLEDGMENT

We thank the referee for carefully reading the manuscript and for improving its presentation by constructive comments.

REFERENCES

1. H. Bauer, "Probability Theory and Elements of Measure Theory," Academic Press, San Diego, 1981.
2. R. G. Casey and E. Lecolinet, A survey of methods and strategies in character segmentation, *IEEE Trans. Pattern Anal. Mach. Intel.* **18** (1996), 690–706.
3. F. A. Graybill, "Matrices with Applications in Statistics," second edition, Wadsworth, Belmont, CA, 1983.
4. K. V. Mardia, J. T. Kent, and J. M. Bibby, "Multivariate Analysis," Academic Press, San Diego, 1979.
5. G. Ritter, Classification and clustering of objects with variants, *in* "Data Analysis" (W. Gaul, O. Opitz, and M. Schader, Eds.), pp. 41–50, 2000.
6. G. Ritter and M. T. Gallegos, A Bayesian approach to object identification in pattern recognition, *in* "Proc. 15th Int. Congress on Pattern Recognition, Barcelona 2000" (A. Sanfeliu *et al.*, Eds.), Vol. 2, pp. 418–421, 2000.
7. G. Ritter and Ch. Pesch, Polarity-free automatic classification of chromosomes, *Comput. Statist. Data Anal.* **35** (2001), 351–372.
8. G. Ritter and G. Schreib, Profile and feature extraction from chromosomes, *in* "Proc. 15th Int. Congress on Pattern Recognition, Barcelona 2000" (A. Sanfeliu *et al.*, Eds.), Vol. 2, pp. 287–290, 2000.
9. G. Ritter and G. Schreib, Using dominant points and variants for profile extraction from chromosomes, *Pattern Recognition* **34** (2001), 157–172.